Replication Timing Networks: a novel class of gene regulatory networks

Juan Carlos Rivera-Mulia[1,†], Sebo Kim[2,†], Haitham Gabr[2], Tamer Kahveci[2,*] and David M. Gilbert[1,3,*]

[1]Department of Biological Science, Florida State University, Tallahassee, FL, 32306-4295, USA.

[1]Department of Computer and Information Sciences and Engineering, University of Florida, Gainesville, Florida 32611, USA.

[3]Center for Genomics and Personalized Medicine, Florida State University, Tallahassee, FL, USA.

†These authors contributed equally to this work.

*Correspondence to: gilbert@bio.fsu.edu and tamer@cise.ufl.edu

Summary

DNA replication occurs in a defined temporal order known as the replication-timing (RT) program and is regulated during development, coordinated with 3D genome organization and transcriptional activity. Here, we exploit genome-wide RT profiles from 15 human cell types and intermediate differentiation stages derived from human embryonic stem cells to construct different types of RT regulatory networks. First, we constructed networks based on the coordinated RT changes during cell fate commitment to create RT networks composed of specific functional sub-network communities. We also constructed directional regulatory networks based on the order of RT changes within cell lineages and identified master regulators of differentiation pathways. Finally, we explored relationships between RT networks and transcriptional regulatory networks (TRNs), by combining them into more complex circuitries of composite and bipartite networks. Our findings show that RT networks can be exploited to dissect the cellular mechanisms that regulate lineage specification and cellular identity maintenance.

Highlights

- DNA replication timing (RT) programs were used to construct gene regulatory networks.
- RT networks revealed functional organization of sub-network communities.
- RT networks identified master regulators of cell fate commitment.
- RT and gene expression circuitries define composite and bipartite networks.

Introduction

During development, specific transcriptional programs and epigenetic landscapes are established that maintain the identities and functionality of the specialized cell types that emerge. Despite characterization of changes in transcriptome and epigenome during development (Gifford et al., 2013; Roadmap Epigenomics Consortium et al., 2015; Tsankov et al., 2015; Xie et al., 2013), little is known about the role of spatio-temporal genome organization in cell fate specification. Changes in gene activity and chromatin 3D organization are coordinated with dynamic changes in the temporal order of genome duplication, known as the replication timing (RT) program (Hiratani et al., 2010; 2008; Rivera-Mulia et al., 2015). Spatio-temporal control of RT is conserved in all eukaryotes (Rivera-Mulia and Gilbert, 2016a; Solovei et al., 2016). RT is regulated during development in discrete chromosome units, referred to as replication domains (RDs), that align with topological associated domains (TADs) mapped by chromosome conformation capture techniques (Hi-C) and segregate into distinct nuclear compartments visualized by either cytogenetic or Hi-C methods (Jackson and Pombo, 1998; Moindrot et al., 2012; Pope et al., 2014; Rivera-Mulia and Gilbert, 2016b; Ryba et al., 2010; Sadoni et al., 2004; Yaffe et al., 2010). Hence, we reasoned that RT can be exploited to characterize regulatory relationships between 3D genome organization and gene expression control during development

Previously, we generated the most comprehensive database of RT programs during human development and found that approximately half of the genome undergoes dynamic changes that are closely coordinated with the establishment of transcriptional programs (Rivera-Mulia et al., 2015). Additionally, we demonstrated that genes within developmentally RT regulated domains are high in the hierarchy of transcriptional regulatory networks (TRNs) and regulate RT constitutive genes (Rivera-Mulia et al., 2015). However, strong gene expression was not restricted to early replicating genomic regions and transcriptional activation during cell fate commitment often preceded RT changes (Rivera-Mulia et al., 2015; Rivera-Mulia and Gilbert, 2016b). In fact, although a long-standing correlation between early replication and gene expression has been observed in all eukaryotes, the link between RT and transcriptional activity is complex and causal relationships have not been established (Rivera-Mulia and Gilbert, 2016b; Solovei et al., 2016). Here, we test the hypothesis that RT can be regulated by the establishment of complex regulatory circuits of transcription factors rather than by the transcription levels of genes within each RD. We extracted RT values at the transcription start

sites (TSS) for all RefSeq genes (Ryba et al., 2011a) and constructed distinct types of RT regulatory network models based on: 1) correlation patterns in RT changes during cell fate commitment, 2) the temporal order of RT changes in each developmental transition and 3) combined networks that explore the crosstalk between RT and transcriptional regulatory networks (composite and bipartite networks).

Results

Construction of RT networks

To construct RT regulatory networks we defined a model that describes the relationship between all possible combinations of gene pairs (nodes), establishing gene interactions (edges) according to their correlated RT patterns during cell differentiation (Figure 1A). Distinct filters were applied in our model: a) we included only genes that change RT during development (removing RT constitutive genes, see methods), b) we established gene interactions (edges) between highly correlated gene pairs (correlation >0.75), c) we included edges only between genes separated by at least 500kb and/or in different chromosomes (Figures 1B-C and S1). Separation by >500kb was chosen to remove gene pairs within the same RD, which we have shown vary in size from 0.4 to 0.8 Mb (Hiratani et al., 2008; Pope et al., 2014; Rivera-Mulia et al., 2015). After applying these filters, gene pairs were extracted (green boxes in Figure 1C) and RT networks were constructed based on the *Pearson's* correlation strength. Figure 1D and E illustrate hypothetical examples of two distinct RT patterns along a single cell differentiation lineage, the correlations for which constitute connections between gene pairs exploited to construct the corresponding RT networks.

RT networks reflect functional gene regulatory interactions

To determine the biological significance of RT networks, we examined their functional organization by performing ontology analysis of each sub-network community using the spatial analysis of functional enrichment (SAFE) algorithm (Baryshnikova, 2016; Costanzo et al., 2016). In order to identify functional sub-network communities, 2D maps of RT networks constructed as described above were generated by force-directed layout algorithm in Cytoscape (Shannon et al., 2003). Next, local neighborhoods within the global networks were identified according to the connectivity and distances between nodes (Blondel et al., 2008) and ontology analysis was performed to identify the most significant enrichment of functional attributes per neighborhood (Baryshnikova, 2016). To test our RT networks models, we first constructed an RT network using all RT correlated gene-pairs across all differentiation pathways (Figure 2A). Interestingly, we found that nodes were arranged in sub-network communities defined by interconnected genes; among those communities we found local neighborhoods with highly interconnected nodes of genes involved in specific functions, which were color coded based on the enrichment of functional ontology annotations (Figure 2A). Closer inspection of local neighborhoods

annotated with specific functions grouped the genes according to their ontology terms (Figure 2B). Next, we generated RT networks independently for ectoderm, mesoderm and endoderm differentiation pathways. Remarkably, these RT networks contained local neighborhoods of genes annotated with specific functions related to each germ layer (Figure 2C-E). Overall, our results revealed that dynamic changes in RT are organized into complex regulatory networks linked to gene functions established during cell fate commitment.

Directional RT networks identify master regulators of cell fate commitment

Interestingly, for each of the RT regulatory networks constructed either for all cell types together (Figure 2A-B) or for each one of the germ layers separately (Figure 2C-E), we identified a local neighborhood associated with transcription factor (TF) activity. These findings suggest that gene regulation by TFs might be critical not only for the establishment of cell type-specific transcriptional programs but also for RT program control. Hence, in order to explore the hierarchical relationships in RT changes during development we constructed directional RT networks for each of the specific differentiation pathways towards pancreas, liver, smooth muscle, mesothelium, mesenchymal stem cells (MSCs) and neural precursors (NPCs). First, we classified the genes according to the order of RT changes during each differentiation pathway (Figure 3A), identified those that change during the earliest cell fate transitions and assigned directional edges to genes that changed in subsequent differentiation stages (see Methods). Directional RT networks were displayed either in 2D maps or in a hierarchical arrangement and nodes were color/size coded according to the order of the changes in RT during distinct differentiation pathways (Figure 3B-C). Constructing of directional RT regulatory networks allowed us to identify the earliest RT changes during development that constitute the *master regulators* of the gene interactions within lineage-specific RT networks (red nodes in Figure 3B-C), as well as their targets and downstream relationships. Consistently, among the genes that change RT during the earliest developmental transitions, we identified key known regulators for specific differentiation pathways such as *SOX17* for liver and pancreas, *SOX1* for mesenchymal stem cells and *MSX2* for smooth muscle, validating these nodes as master regulators of cell fate commitment (Figures 3D-E and S2).

To characterize the distinct levels of regulation we characterized the distribution of nodes in each hierarchy level. *Master regulators* were defined as all genes that change RT in the earliest differentiation transition (red node in Figure 3), while downstream nodes were classified according to the time during differentiation when they change RT: either as *managers* or *effectors* (green and blue nodes and grey nodes respectively in Figure 3F). Interestingly, we found that developmental establishment of RT networks occurs differently for each germ layer. For endoderm cell types (liver and pancreas) most of the changes occur very early during differentiation, with many master regulators changing to early replication and fewer downstream nodes in each of the subsequent differentiation stages (Figure 3G). In contrast, for mesoderm cell types (smooth muscle and mesothelium) few master regulators were connected with many downstream nodes. These differences may reflect fundamental principles regulating germ layer specification. Next, to identify the classes of genes in each hierarchical level, we performed an ontology analysis (Mi et al., 2017) and found that master regulators are enriched in receptor binding genes (growth factor receptors) and TFs (Figure 3H). This analysis reveals an intimate relationship between RT regulation and key regulators of cellular differentiation.

RT network edges overlap with known transcriptional regulatory interactions

Since we found that transcription factors are among the master regulators of RT networks, we analyzed whether the interactions within RT networks overlap with known gene regulatory interactions in transcriptional regulatory networks (TRNs) using a previously described set of cell type-specific networks of TFs (Neph et al., 2012). First, we identified the cell types that most closely match the TRNs to our RT networks, as follows: hESC-derived hepatocytes were compared to TRNs from HepG2 – a liver cancer cell line that retains morphological and functional hepatocyte properties (Berger et al., 2015; Knowles et al., 1980), hESC-derived mesothelial cells were compared to TRNs from HCF cells – cardiac fibroblasts that during development and *in vitro* differentiation can be derived from mesothelial cells (Mutsaers, 2004) and hESC-derived neural precursors were compared to TRNs from the SK-N-SH cell line after treatment with retinoic acid – SK-N-SH cells were derived from a neuroblastoma and retinoic acid causes differentiation to neural phenotype (Preis et al., 1988). Next, we constructed RT networks using only the subset of genes present in the TRNs (475 TFs) that change RT and are highly RT-correlated in each differentiation pathway (Pearson's >0.75). Finally, we identified the number of common and unique edges to RT networks and TRNs. We found that in all three cases there was considerable and highly significant overlap when compared to the expected

overlap by randomly selecting the same number of edges (Figure 4A). In fact, significant overlap was also observed when all cell types from both RT networks and TRNs were classified per germ layer (Table S1) and common edges were identified for ectoderm and mesoderm, even when distinct cell types were used for each germ layer. These results further validate the gene regulatory interactions amongst genes identified in RT networks.

Building blocks of RT networks are small motifs of >4 nodes that lack TRN edges

Previous studies have explored the architecture of gene regulatory relationships by analyzing either transcriptional or protein interactions and found that complex cellular networks are constituted by sets of small network motifs, such as interactions between transcription factors and their targets (Alon, 2007; Zhang et al., 2005). Here, we performed a topology characterization of RT networks constructed with the subset of genes present in the TRNs –475 TFs (Neph et al., 2012), to explore the most significant patterns of interactions. We computed all possible motifs composed by 2-4 nodes and identified the motifs with high occurrence in each RT network constructed per differentiation pathway (Figures S3-S5). Statistical significance of each motif pattern was calculated by comparison to randomized networks (Baiser et al., 2015; Elhesha and Kahveci, 2016; Milo et al., 2002). Consistent with previous observations in transcriptional and protein regulatory networks, we found that the most frequent motifs consist of interactions between less than 4 nodes for all differentiation pathways analyzed (Figure 4B). Similar results were observed when motifs were identified from all RT correlated genes in all cell types within each germ layer (endoderm, ectoderm and mesoderm). Since the predominant building blocks are small motifs with few nodes, RT networks of TFs are composed of separated multiple sub-networks. Additionally, since significant overlap between RT and transcriptional networks was observed in all differentiation pathways, we examined the presence of transcriptional edges within the RT networks motifs. Surprisingly, we found that although distinct transcriptional edges were present within some of the RT network motifs, the most frequent motifs did not contain TRN edges (Figures S3-S5), consistent with our hypothesis that RT can be regulated by the establishment of complex regulatory circuits of transcription factors rather than by the transcription levels of genes within each RD.

Composite networks: combining RT and transcriptional regulatory networks

Intriguingly, although our findings demonstrate that RT networks are linked to cellular function and differentiation processes and significantly overlap with well characterized TRNs of TFs, our results above demonstrate that the majority of the RT networks motifs do not contain transcriptional edges. Hence, to better understand how the regulatory circuitries are established during cell fate commitment we constructed a model of composite regulatory networks by merging RT and TRNs. Previous studies have demonstrated that distinct types of interactions (such as protein-protein and transcription regulation) can be combined to explore more complex cellular circuitries (Vidal et al., 2011; Yeger-Lotem et al., 2004). Here, we combined the interactions observed in RT networks with those in the TRNs between TFs. First, we used as base networks the set of motifs from each RT network and identified the nodes that are also connected in TRNs (Figure 4C). Then we identified all the interactions between those RT nodes in the TRNs from matching cell types and constructed composite networks by adding the transcriptional interactions (Figure 4C). Interestingly, RT networks for each differentiation pathway are constituted by multiple unconnected motifs of >4 nodes; however, the addition of transcriptional edges revealed more complex and highly interconnected networks with all nodes interacting with at least 3 other nodes (Figure 4B-D). Importantly, every composite network contained known key regulators for each differentiation pathway such as SOX17 and GLIS3 for liver and MSX2, FOXP1 and WT1 for mesothelium (Figure 4D-E). Hence, our composite networks reveal biologically relevant gene interactions important for cell fate commitment.

Bipartite networks reveal transcription factors as regulators of RT

To further explore the relationship between RT and gene expression we identified the most significant genes expressed in each cell type and analyzed their correlation to RT networks. First, we analyzed genome-wide transcriptomes for the same cell types from which we obtained the RT programs (Rivera-Mulia et al., 2015). Our highly comprehensive characterization of gene expression, including multiple replicates for each differentiation stage, allowed us to identify with confidence the genes that are differentially expressed during cell fate commitment towards each cell type and the genes that better distinguish each intermediate stage. Co-expressed genes were identified by weighted correlation network analysis (Langfelder and Horvath, 2008). Strong correlations between gene expression levels are widely

used to identify regulatory interactions (Allocco et al., 2004; D'haeseleer et al., 2000; Gabr et al., 2015; Horvath and Dong, 2008; Laurenti et al., 2013; Li, 2002; Novak and Jain, 2006); thus, we used the most significant co-expressed genes to construct TRNs for each differentiation pathway. To decrease the complexity of the data we focused in the top 100 genes that are significantly co-expressed in specific cell types/intermediate differentiation stages. In all differentiation pathways and for each differentiation stage we found that transcription factors were among the most significant genes distinguishing each cell type (Figure 5A). Moreover, ontology analysis (Ashburner et al., 2000; The Gene Ontology Consortium, 2015) using the different subsets of genes revealed strong enrichment of genes regulating the development of each cell type (Table S2).

Since we found that: a) TFs are among the master regulators in RT networks, b) interactions between TFs in TRNs significantly overlap with RT networks, c) key TFs are among the highest interconnected nodes in RT networks and, d) TF expression patterns distinguish each cell type, we next addressed whether TFs might be involved in the establishment of RT programs during development, and whether that role might be independent of their role in regulating transcription. Thus, we identified the genes whose RT patterns are highly correlated with the expression levels of the top 100 genes that distinguish each cell type. Then, we employed the correlation patterns between gene expression and RT changes to construct bipartite networks. Exemplary gene expression levels from a subset of TFs critical for pancreas development is shown in Figure 5B as well as genes with correlated patterns of RT regulation during pancreatic differentiation. Bipartite networks consist of two independent but interconnected networks: the TRN side contains genes that are co-expressed in specific developmental stages/cell types and the RT side contains genes whose RT changes were highly correlated with the expression patterns from the TRN side. An exemplary bipartite network for pancreatic development is shown in Figure 5C. Remarkably, ontology analysis using the set of genes of each sides of the bipartite network resulted in specific annotations relevant for regulation of pancreatic differentiation (Figure 5C). The bipartite network shown in Figure 5C was constructed using only the TFs co-expressed in pancreatic cells for visualization purposes; however, similar results and ontology terms were found using the complete set of genes co-regulated during pancreatic differentiation. These results further support our hypothesis that TRNs influence replication timing by mechanisms that are separated by the transcriptional regulation of downstream targets of TFs.

Discussion

In this study, we introduced a new approach to construct gene regulatory networks exploiting the dynamic changes in DNA replication timing during development. RT is cell type-specific (Hiratani et al., 2010; Rivera-Mulia et al., 2015; Ryba et al., 2011b), regulation of RT is critical to maintain genome stability (Alver et al., 2014; Donley et al., 2013; Neelsen et al., 2013) and abnormal RT is observed in disease (Dixon et al., 2017; Gerhardt et al., 2014a; 2014b; Ryba et al., 2012; Sasaki et al., 2017). Moreover, RT is closely related to the spatio-temporal organization of the genome with early and late replicating domains segregating to distinct nuclear compartments (Pope et al., 2014; Rivera-Mulia and Gilbert, 2016b). Furthermore, cell fate commitment is accompanied by dynamic changes in RT that are globally coordinated with transcriptional activity (Rivera-Mulia et al., 2015; Rivera-Mulia and Gilbert, 2016b). Hence, RT constitutes a functional readout of genome organization that is linked to gene regulation during development. We constructed RT networks based on the RT changes across 15 cell types and differentiation intermediates derived from human embryonic stem cells. We identified thousands of genes from different chromosomes that are co-regulated in RT during development (Figure 1C) and constructed distinct RT network models based on their dynamic changes.

Confirming the link between RT and gene regulation, RT networks constructed based on correlated changes in RT are organized into multiple sub-network communities with neighborhoods associated with specific functions (Figure 2). We also developed a model of directional RT networks able to explore the hierarchical relationships between RT co-regulated genes and identify the master regulators of cell differentiation and their downstream targets. To validate our model of directional RT networks we characterized the gene interactions of specific differentiation pathways and consistently obtained key regulators of cell fate commitment as master regulators (Figure 3). The algorithms to construct these RT networks can be applied to explore the interactions of any gene of interest (see Methods for detailed information on the computational pipeline).

Combination of TRNs and RT networks into composite and bipartite networks revealed new insights into gene regulation during cell fate commitment. First, we found that there is a highly significant overlap between TRNs and RT networks of TFs (Figure 4A). However, the RT

network motifs did not contain transcriptional edges, suggesting that the RT co-regulated TFs do not regulate each other's transcription levels during cell fate commitment. Hence, establishment of RT regulatory networks of specific subsets of TFs might be necessary for the proper regulation (either at the level of RT or gene expression) of downstream genes in a more complex regulatory circuit to ensure and maintain the distinct cell identities. For example, TRNs may regulate RT independent of their direct role in transcriptional regulation, which in turn affects the responsiveness of RT-regulated genes to downstream transcriptional regulation. Second, composite networks solved the conundrum of high overlap between RT and TRNs with a lack of transcriptional interactions within the RT networks motifs; composite networks revealed more complex circuitries in which transcriptional edges connected otherwise separated RT motifs (Figure 4D-E). Finally, construction of bipartite networks confirmed the key role of transcription factors in regulation of both transcriptional and RT programs during cell fate commitment (Figure 5), with hundreds of RT co-regulated genes correlated with expression levels of co-expressed TFs within the same differentiation pathways. This is of particular significance to our understanding of how RT is related to cell fate changes because, despite the correlation between early replication and transcriptional activity, no causal links have been unveiled (Rivera-Mulia and Gilbert, 2016b). In fact, knockout/ knockdown or overexpression of many transcription and chromatin structure regulators (including TFs such as C-MYC, N-MYC, MYOD and PAX5) has no effect on RT (Dileep et al., 2015) and combinatorial co-regulation of multiple TFs might be required to control TRNs (Gerstein et al., 2012; Novershtern et al., 2011). Hence, establishment of complex circuitries/complete regulatory TFs networks, rather than transcriptional induction of specific downstream targets, might be required to shape the RT program during development.

Author Contributions

JCRM and SK contributed equally to this work; JCRM, TK and DMG conceived and designed the study; JCRM, SK and HG performed data analysis and interpretation; JCRM and DMG wrote the manuscript.

Acknowledgments

This work was supported by NIH grant GM083337 and GM085354 (D.M.G.).

References

Allocco, D.J., Kohane, I.S., Butte, A.J., 2004. Quantifying the relationship between co-expression, co-regulation and gene function. BMC Bioinformatics 5, 18. doi:10.1186/1471-2105-5-18

Alon, U., 2007. Network motifs: theory and experimental approaches. Nat Rev Genet 8, 450–461. doi:10.1038/nrg2102

Alver, R.C., Chadha, G.S., Blow, J.J., 2014. The contribution of dormant origins to genome stability: from cell biology to human genetics. DNA Repair (Amst.) 19, 182–189. doi:10.1016/j.dnarep.2014.03.012

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29.

Baiser, B., Elhesha, R., Kahveci, T., 2015. Motifs in the assembly of food web networks. Oikos 125, 480–491. doi:10.1111/oik.02532

Baryshnikova, A., 2016. Systematic Functional Annotation and Visualization of Biological Networks. Cell Syst 2, 412–421. doi:10.1016/j.cels.2016.04.014

Berger, E., Vega, N., Weiss-Gayet, M.X., le, xe9, G., xeb, lo, n, A., 2015. Gene Network Analysis of Glucose Linked Signaling Pathways and Their Role in Human Hepatocellular Carcinoma Cell Growth and Survival in HuH7 and HepG2 Cell Lines. BioMed Research International 2015, 1–19. doi:10.1155/2015/821761

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech. 2008, P10008. doi:10.1088/1742-5468/2008/10/P10008

Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., Pelechano, V., Styles, E.B., Billmann, M., van Leeuwen, J., van Dyk, N., Lin, Z.-Y., Kuzmin, E., Nelson, J., Piotrowski, J.S., Srikumar, T., Bahr, S., Chen, Y., Deshpande, R., Kurat, C.F., Li, S.C., Li, Z., Usaj, M.M., Okada, H., Pascoe, N., Luis, B.-J.S., Sharifpoor, S., Shuteriqi, E., Simpkins, S.W., Snider, J., Suresh, H.G., Tan, Y., Zhu, H., Malod-Dognin, N., Janjic, V., Przulj, N., Troyanskaya, O.G., Stagljar, I., Xia, T., Ohya, Y., Gingras, A.-C., Raught, B., Boutros, M., Steinmetz, L.M., Moore, C.L., Rosebrock, A.P., Caudy, A.A., Myers, C.L., Andrews, B., Boone, C., 2016. A global genetic interaction network maps a wiring diagram of cellular function. Science 353, aaf1420–aaf1420. doi:10.1126/science.aaf1420

Dileep, V., Rivera-Mulia, J.C., Sima, J., Gilbert, D.M., 2015. Large-Scale Chromatin Structure-Function Relationships during the Cell Cycle and Development: Insights from Replication Timing. Cold Spring Harb. Symp. Quant. Biol. 80, 53–63. doi:10.1101/sqb.2015.80.027284

Dixon, J., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardimci, G.G., Chakraborty, A., Bann, D.V., Wang, Y., Clark, R., Zhang, L., Yang, H., Liu, T., Iyyanki, S., An, L., Pool, C., Sasaki, T., Rivera-Mulia, J.C., Ozadam, H., Lajoie, B.R., Kaul, R., Buckley, M., Lee, K., Diegel, M., Pezic, D., Ernst, C., Hadjur, S., Odom, D.T., Stamatoyannopoulos, J.A., Broach, J.R., Hardison, R., Ay, F., Noble, W.S., Dekker, J., Gilbert, D.M., Yue, F., 2017. An Integrative Framework For Detecting Structural Variations In Cancer Genomes. doi:10.1101/119651

Donley, N., Stoffregen, E.P., Smith, L., Montagna, C., Thayer, M.J., 2013. Asynchronous Replication, Mono-Allelic Expression, and Long Range Cis-Effects of ASAR6. PLoS Genet. 9, e1003423. doi:10.1371/journal.pgen.1003423

D'haeseleer, P., Liang, S., Somogyi, R., 2000. Genetic network inference: from co-expression clustering to reverse engineering. Bioinformatics 16, 707–726. doi:10.1093/bioinformatics/16.8.707

Elhesha, R., Kahveci, T., 2016. Identification of large disjoint motifs in biological networks. BMC Bioinformatics 17, 408. doi:10.1186/s12859-016-1271-7

Gabr, H., Rivera-Mulia, J.C., Gilbert, D.M., Kahveci, T., 2015. Computing interaction probabilities in signaling networks. EURASIP J Bioinform Syst Biol 2015, 10. doi:10.1186/s13637-015-0031-8

Gerhardt, J., Tomishima, M.J., Zaninovic, N., Colak, D., Yan, Z., Zhan, Q., Rosenwaks, Z., Jaffrey, S.R., Schildkraut, C.L., 2014a. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. Mol. Cell 53, 19–31. doi:10.1016/j.molcel.2013.10.029

Gerhardt, J., Zaninovic, N., Zhan, Q., Madireddy, A., Nolin, S.L., Ersalesi, N., Yan, Z., Rosenwaks, Z., Schildkraut, C.L., 2014b. Cis-acting DNA sequence at a replication origin promotes repeat expansion

to fragile X full mutation. The Journal of Cell Biology 206, 599–607. doi:10.1083/jcb.201404157

Gerstein, M.B., Kundaje, A., Hariharan, M., Landt, S.G., Yan, K.-K., Cheng, C., Mu, X.J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman, N., Bhardwaj, N., Boyle, A.P., Cayting, P., Charos, A., Chen, D.Z., Cheng, Y., Clarke, D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A., Jain, P., Kasowski, M., Lacroute, P., Leng, J., Lian, J., Monahan, H., O'Geen, H., Ouyang, Z., Partridge, E.C., Patacsil, D., Pauli, F., Raha, D., Ramirez, L., Reddy, T.E., Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K.Y., Zilberman-Schapira, G., Batzoglou, S., Sidow, A., Farnham, P.J., Myers, R.M., Weissman, S.M., Snyder, M., 2012. Architecture of the human regulatory network derived from ENCODE data. Nature 489, 91–100. doi:doi:10.1038/nature11245

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., Zhang, X., Coyne, M., Fostel, J.L., Holmes, L., Meldrim, J., Guttman, M., Epstein, C., Park, H., Kohlbacher, O., Rinn, J., Gnirke, A., Lander, E.S., Bernstein, B.E., Meissner, A., 2013. Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. Cell 153, 1149–1163. doi:10.1016/j.cell.2013.04.037

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., Rathjen, P.D., Gilbert, D.M., 2010. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. 20, 155–169. doi:10.1101/gr.099796.109

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T.M., Schübeler, D., Gilbert, D.M., 2008. Global reorganization of replication domains during embryonic stem cell differentiation. PLoS Biol. 6, e245. doi:10.1371/journal.pbio.0060245

Horvath, S., Dong, J., 2008. Geometric Interpretation of Gene Coexpression Network Analysis. PLoS Comput. Biol. 4, e1000117. doi:10.1371/journal.pcbi.1000117

Jackson, D.A., Pombo, A., 1998. Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells. The Journal of Cell Biology 140, 1285–1295. doi:10.1083/jcb.140.6.1285

Knowles, B.B., Howe, C.C., Aden, D.P., 1980. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science 209, 497–499. doi:10.1126/science.6248960

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559. doi:10.1186/1471-2105-9-559

Laurenti, E., doulatov, S., Zandi, S., Plumb, I., Chen, J., April, C., Fan, J.-B., Dick, J.E., 2013. The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment. Nat Immunol 14, 756–763. doi:10.1038/ni.2615

Li, K.-C., 2002. Genome-wide coexpression dynamics: theory and application. Proc. Natl. Acad. Sci. U.S.A. 99, 16875–16880. doi:10.1073/pnas.252466999

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D., 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189. doi:10.1093/nar/gkw1138

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U., 2002. Network Motifs: Simple Building Blocks of Complex Networks. Science 298, 824–827. doi:10.1126/science.298.5594.824

Moindrot, B., Audit, B., Klous, P., Baker, A., Thermes, C., de Laat, W., Bouvet, P., Mongelard, F., Arneodo, A., 2012. 3D chromatin conformation correlates with replication timing and is conserved in resting cells. 40, 9470–9481. doi:10.1093/nar/gks736

Mutsaers, S.E., 2004. The mesothelial cell. The International Journal of Biochemistry & Cell Biology 36, 9–16. doi:10.1016/S1357-2725(03)00242-5

Neelsen, K.J., Zanini, I.M.Y., Mijic, S., Herrador, R., Zellweger, R., Ray Chaudhuri, A., Creavin, K.D., Blow, J.J., Lopes, M., 2013. Deregulated origin licensing leads to chromosomal breaks by rereplication of a gapped DNA template. Genes & Development 27, 2537–2542. doi:10.1101/gad.226373.113

Neph, S.S., Stergachis, A.B.A., Reynolds, A.A., Sandstrom, R.R., Borenstein, E.E., Stamatoyannopoulos, J.A.J., 2012. Circuitry and dynamics of human transcription factor regulatory networks. Cell 150, 1274–1286. doi:10.1016/j.cell.2012.04.040

Novak, B.A., Jain, A.N., 2006. Pathway recognition and augmentation by computational analysis of

microarray expression data. Bioinformatics 22, 233–241. doi:10.1093/bioinformatics/bti764

Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., Frampton, G.M., Drake, A.C.B., Leskov, I., Nilsson, B., Preffer, F., Dombkowski, D., Evans, J.W., Liefeld, T., Smutko, J.S., Chen, J., Friedman, N., Young, R.A., Golub, T.R., Regev, A., Ebert, B.L., 2011. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell 144, 296–309. doi:10.1016/j.cell.2011.01.004

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., Thurman, R.E., Cheng, Y., Gülsoy, G., Dennis, J.H., Snyder, M.P., Stamatoyannopoulos, J.A., Taylor, J., Hardison, R.C., Kahveci, T., Ren, B., Gilbert, D.M., 2014. Topologically associating domains are stable units of replication-timing regulation. Nature 515, 402–405. doi:10.1038/nature13986

Preis, P.N., Saya, H., Nádasdi, L., Hochhaus, G., Levin, V., Sadée, W., 1988. Neuronal cell differentiation of human neuroblastoma cells by retinoic acid plus herbimycin A. Cancer Res. 48, 6530–6534.

Rivera-Mulia, J.C., Buckley, Q., Sasaki, T., Zimmerman, J., Didier, R.A., Nazor, K., Loring, J.F., Lian, Z., Weissman, S., Robins, A.J., Schulz, T.C., Menendez, L., Kulik, M.J., Dalton, S., Gabr, H., Kahveci, T., Gilbert, D.M., 2015. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. Genome Res. 25, 1091–1103. doi:10.1101/gr.187989.114

Rivera-Mulia, J.C., Gilbert, D.M., 2016a. Replicating Large Genomes: Divide and Conquer. Mol. Cell 62, 756–765. doi:10.1016/j.molcel.2016.05.007

Rivera-Mulia, J.C., Gilbert, D.M., 2016b. Replication timing and transcriptional control: beyond cause and effect-part III. Curr. Opin. Cell Biol. 40, 168–178. doi:10.1016/j.ceb.2016.03.022

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlić, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. doi:10.1038/nature14248

Ryba, T., Battaglia, D., Chang, B.H., Shirley, J.W., Buckley, Q., Pope, B.D., Devidas, M., Druker, B.J., Gilbert, D.M., 2012. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. Genome Res. 22, 1833–1844. doi:10.1101/gr.138511.112

Ryba, T., Battaglia, D., Pope, B.D., Hiratani, I., Gilbert, D.M., 2011a. Genome-scale analysis of replication timing: from bench to bioinformatics. Nat Protoc 6, 870–895. doi:10.1038/nprot.2011.328

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., Gilbert, D.M., 2010. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. 20, 761–770. doi:10.1101/gr.099655.109

Ryba, T., Hiratani, I., Sasaki, T., Battaglia, D., Kulik, M., Zhang, J., Dalton, S., Gilbert, D.M., 2011b. Replication timing: a fingerprint for cell identity and pluripotency. PLoS Comput. Biol. 7, e1002225. doi:10.1371/journal.pcbi.1002225

Sadoni, N., Cardoso, M.C., Stelzer, E.H.K., Leonhardt, H., Zink, D., 2004. Stable chromosomal units determine the spatial and temporal organization of DNA replication. J. Cell. Sci. 117, 5353–5365. doi:10.1242/jcs.01412

Sasaki, T., Rivera-Mulia, J.C., Vera, D., Zimmerman, J., Das, S., Padget, M., Nakamichi, N., Chang, B.H., Tyner, J., Druker, B.J., Weng, A.P., Civin, C.I., Eaves, C.J., Gilbert, D.M., 2017. Stability of patient-specific features of altered DNA replication timing in xenografts of primary human acute lymphoblastic

leukemia. Exp. Hematol. doi:10.1016/j.exphem.2017.04.004

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. doi:10.1101/gr.1239303

Solovei, I., Thanisch, K., Feodorova, Y., 2016. How to rule the nucleus: divide et impera. Curr. Opin. Cell Biol. 40, 47–59. doi:10.1016/j.ceb.2016.02.014

The Gene Ontology Consortium, 2015. Gene Ontology Consortium: going forward. Nucleic Acids Res. 43, D1049–D1056. doi:10.1093/nar/gku1179

Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., Meissner, A., 2015. Transcription factor binding dynamics during human ES cell differentiation. Nature 518, 344–349. doi:10.1038/nature14233

Vidal, M., Cusick, M.E., Barabási, A.-L., 2011. Interactome Networks and Human Disease. Cell 144, 986–998.

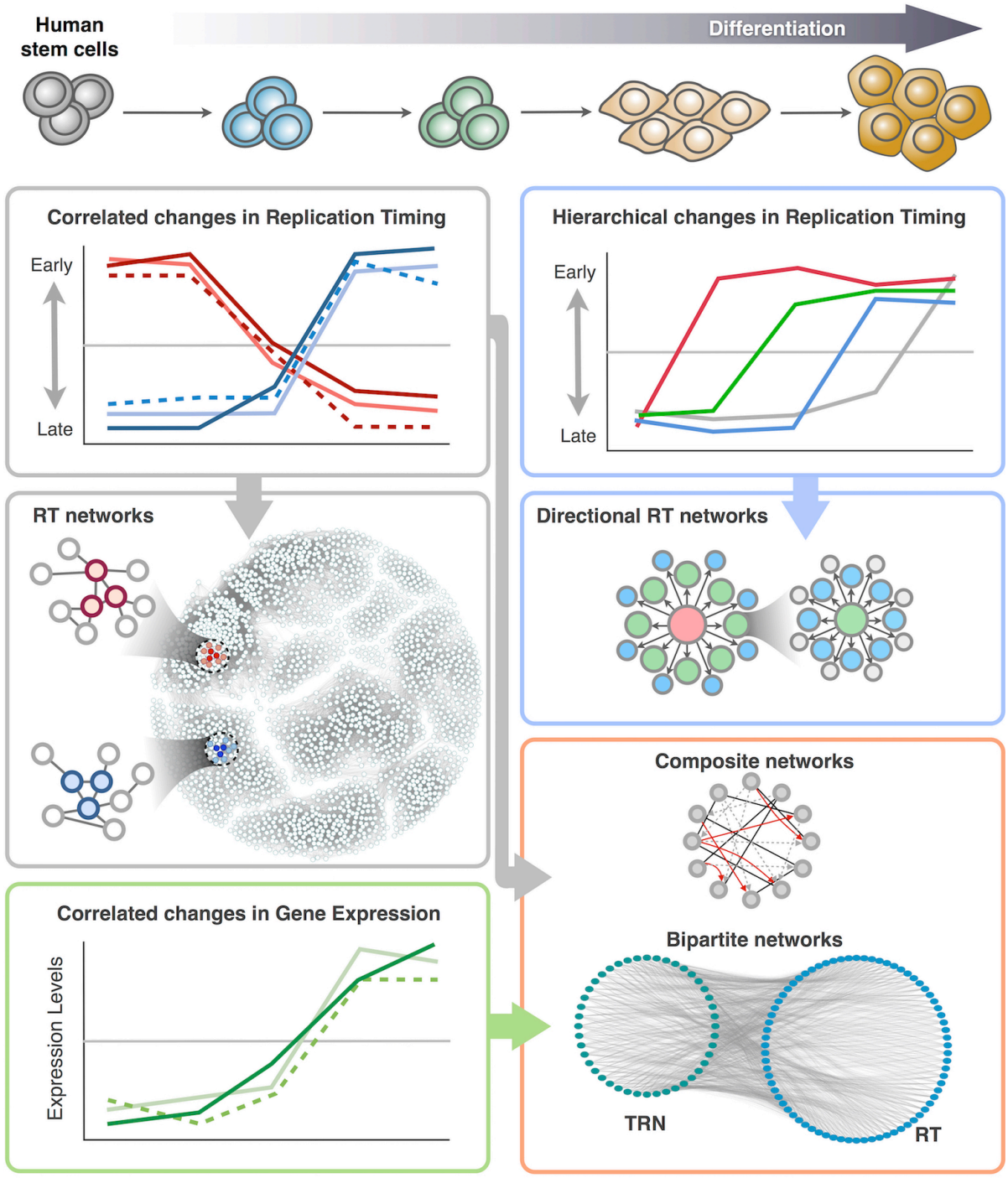
Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., Yang, H., Wang, T., Lee, A.Y., Swanson, S.A., Zhang, J., Zhu, Y., Kim, A., Nery, J.R., Urich, M.A., Kuan, S., Yen, C.-A., Klugman, S., Yu, P., Suknuntha, K., Propson, N.E., Chen, H., Edsall, L.E., Wagner, U., Li, Y., Ye, Z., Kulkarni, A., Xuan, Z., Chung, W.-Y., Chi, N.C., Antosiewicz-Bourget, J.E., Slukvin, I., Stewart, R., Zhang, M.Q., Wang, W., Thomson, J.A., Ecker, J.R., Ren, B., 2013. Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. Cell 153, 1134–1148. doi:10.1016/j.cell.2013.04.022

Yaffe, E., Farkash-Amar, S., Polten, A., Yakhini, Z., Tanay, A., Simon, I., 2010. Comparative Analysis of DNA Replication Timing Reveals Conserved Large-Scale Chromosomal Architecture. PLoS Genet. 6, e1001011. doi:10.1371/journal.pgen.1001011.g005

Yeger-Lotem, E., Sattath, S., Kashtan, N., Itzkovitz, S., Milo, R., Pinter, R.Y., Alon, U., Margalit, H., 2004. Network motifs in integrated cellular networks of transcription-regulation and protein-protein interaction. Proc. Natl. Acad. Sci. U.S.A. 101, 5934–5939. doi:10.1073/pnas.0306752101

Zhang, L.V., King, O.D., Wong, S.L., Goldberg, D.S., Tong, A.H.Y., Lesage, G., Andrews, B., Bussey, H., Boone, C., Roth, F.P., 2005. Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network. J Biol 4, 6. doi:10.1186/jbiol23

510 # Figures



Graphical Abstract

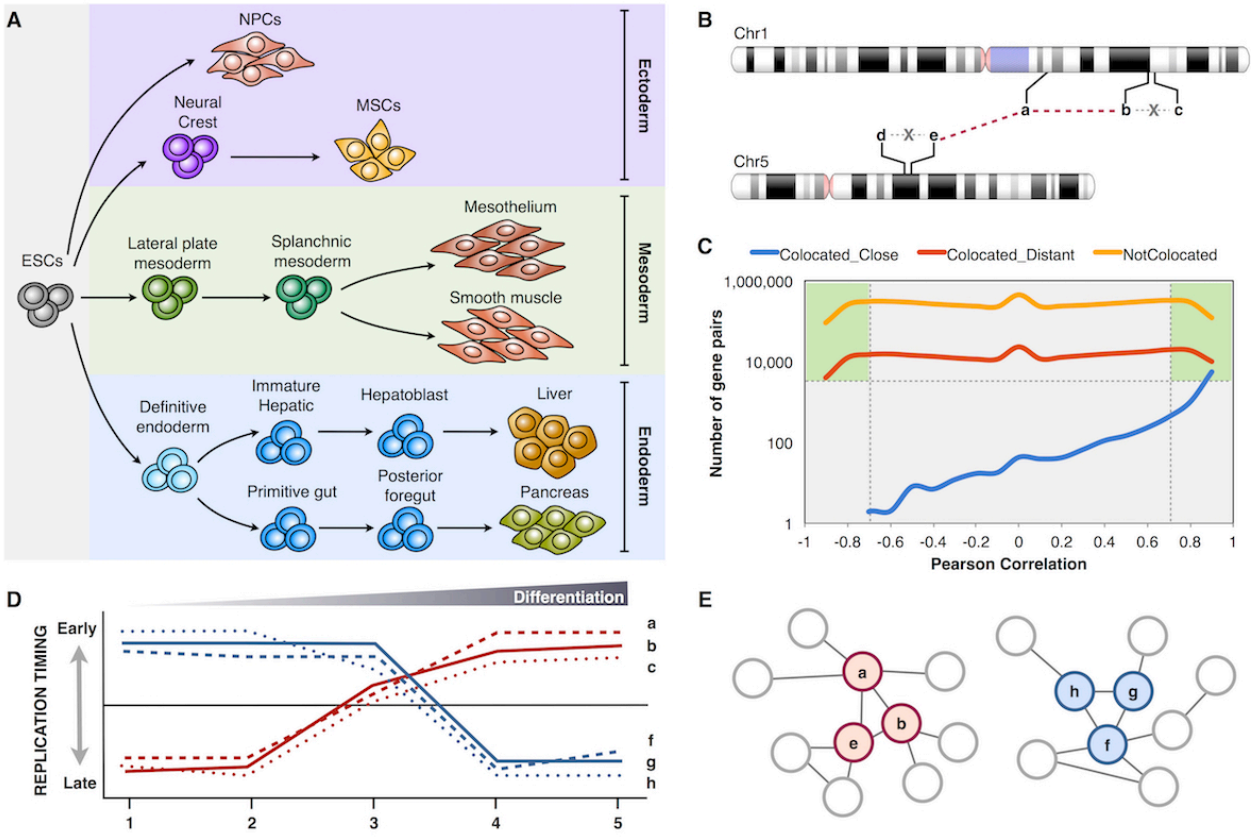


Figure 1. Coordinated changes in RT can be exploited to construct gene regulatory networks.

A) RT programs of distinct cell types representing intermediate stages of human embryonic stem cell differentiation towards endoderm, mesoderm, ectoderm were analyzed for the construction of RT regulatory networks. B) Depiction of different highly RT correlated genes from distinct chromosomes and the establishment of network interaction edges between them. From all possible combinations of gene pairs, those co-located within 500kb were removed from the analysis. Regulatory interactions (edges) between gene pairs are considered only for genes located >500kb apart (co-located distant) or in distinct chromosomes (not co-located), i.e. edges between genes *b-c* and *d-e* were not included in the analysis. C) Number of gene pairs as function of RT correlation for distinct categories of gene pairs: co-located close (within 500kb), co-located distant (separated by > 500kb) and not co-located (from different chromosomes). Only gene pairs with RT correlations >0.75 and located at least 500kb apart were considered. D) RT patterns during cell fate commitment of distinct hypothetical genes. E) Construction of RT regulatory networks based on the *Pearson's* correlation (distance between nodes are proportional to the correlation strength).

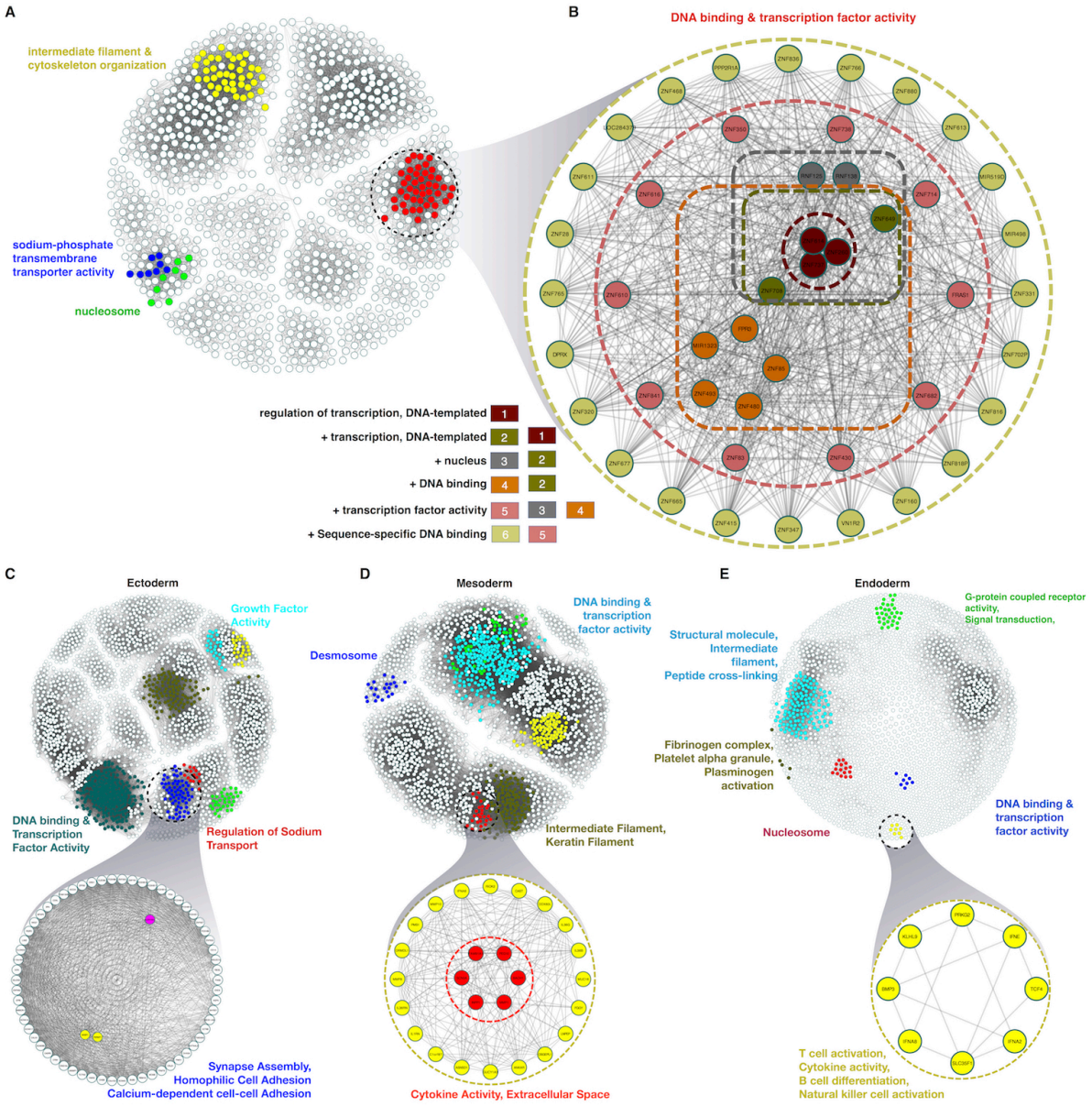


Figure 2. Functional annotation of RT networks.

A) Global RT network constructed from the gene pairs correlated across all differentiation pathways from human ES cells. B) Detailed RT network neighborhood and its respective node organization. C-E) Global RT networks and functional local neighborhoods were constructed for differentiation pathways towards ectoderm (C), mesoderm (D) and endoderm (E). Interaction edges between gene pairs were established for highly correlated nodes (correlation >0.75) and the subset of most connected nodes (>20 edges) were used to visualize RT networks displayed as 2D maps by force-directed network layout algorithm in Cytoscape (Shannon et al., 2003). Highly connected local neighborhoods were annotated with functional ontology terms using SAFE algorithm (Baryshnikova, 2016) and displayed in distinct colors. Specific local neighborhoods were expanded and their nodes were arranged according to their associated GO terms in Cytoscape (Shannon et al., 2003).

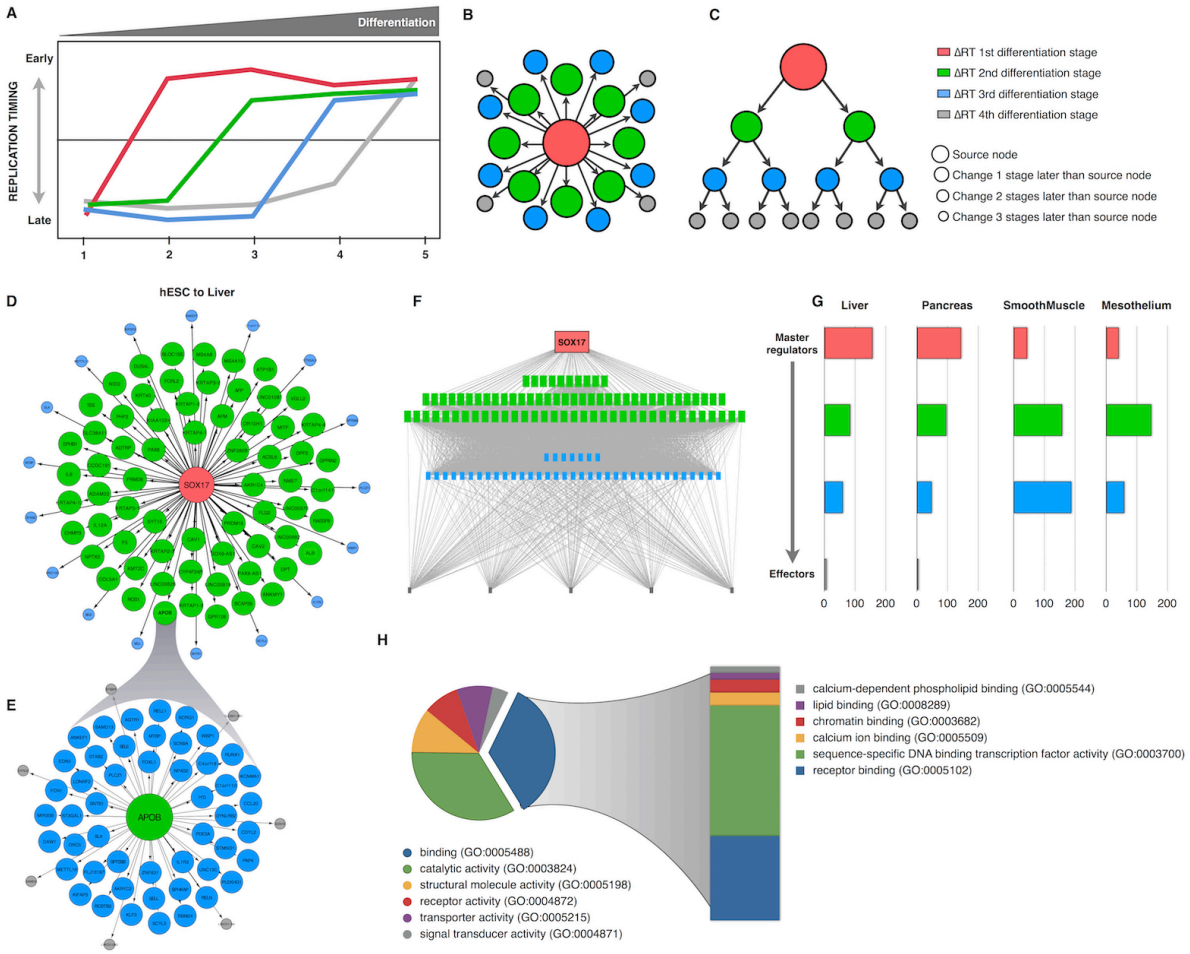


Figure 3: Directional RT networks

A) Distinct genes change RT at different time points during cell fate commitment and the order of RT changes can be used to construct directional RT networks. Red genes change during the first transition between differentiation stages while grey genes change at the last differentiation transition. B) 2D map of a directional RT network were constructed based on a source gene (central node) and downstream connected nodes. C) hierarchical display of RT networks was also constructed based on the order of RT changes. D) An exemplary directional RT network for liver differentiation is shown. The central node is *SOX17* and all downstream nodes were connected based on temporal times during differentiation at which they change RT. E) A *SOX17* downstream network based on the APOB gene exemplify downstream relationships in RT networks. F) *SOX17* is a *master regulator* of endoderm development. The complete hierarchical RT directed network for *SOX17* was constructed including all connected nodes in downstream differentiation stages. Green and blue nodes represent “manager” nodes that are connected to the final “effector” nodes at the lowest level of the network (grey nodes). G) Node distribution in each of the hierarchical levels for each differentiation pathway. H) Ontology analysis of master regulator nodes reveals receptor activation and transcription factor activity as key regulators of RT networks. Ontology analysis classification of gene classes and was performed using protein annotation through evolutionary relationship database –PANTHER (Mi et al., 2017).

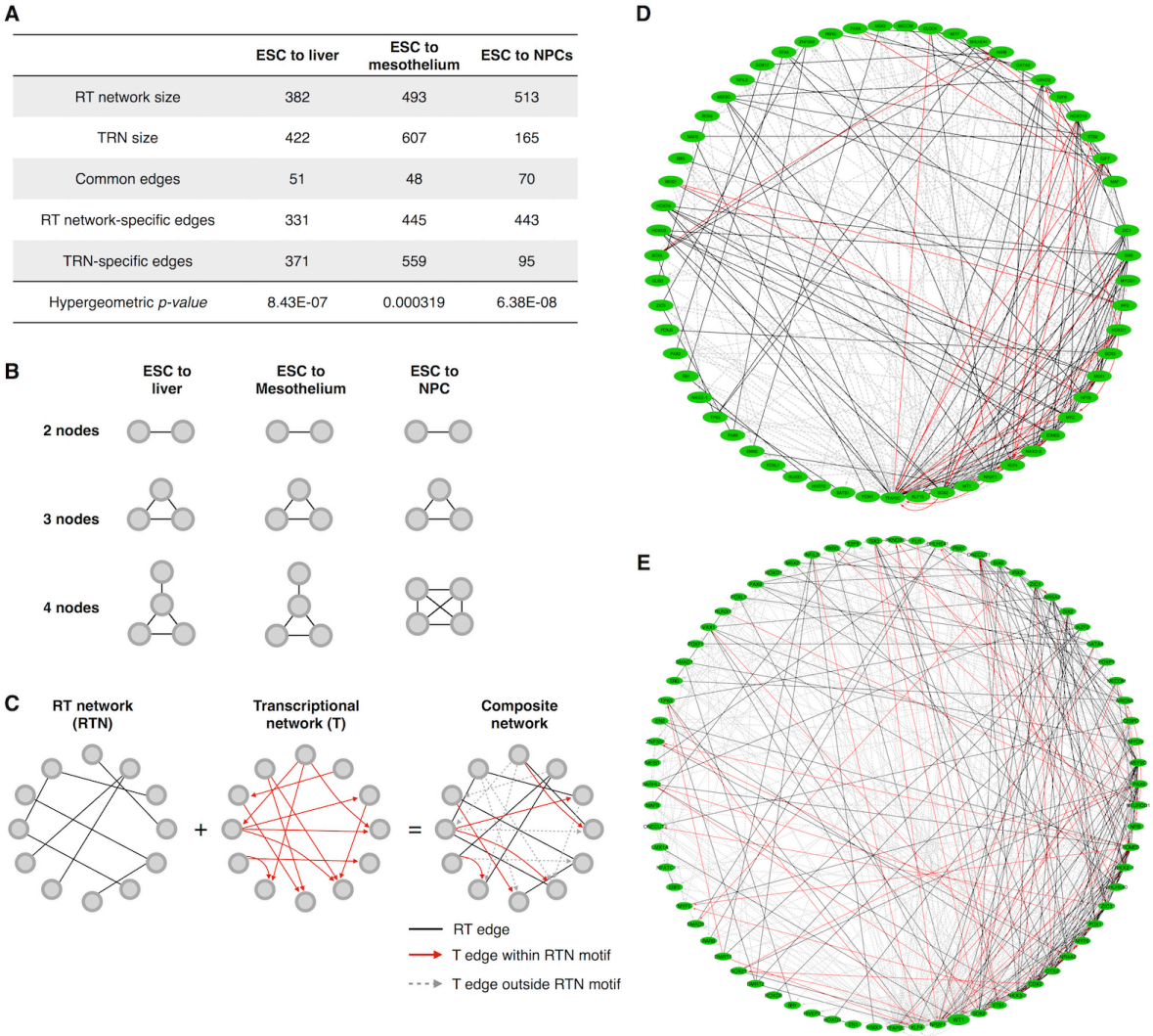


Figure 4: RT and transcriptional networks overlap and can be combined into composite networks.

A) Overlap analysis of RT and TRNs interaction edges. RT networks were constructed for matching cell types in the TRNs (Neph, et al., 2013) and common and unique interaction edges were identified. Only genes within the TRNs were used (475 transcription factors). Hypergeometric test was performed to test the overlap significance (*p-values* are shown). B) Motif analysis revealed that RT networks are constituted by small building blocks (>4 nodes). The most recurrent motif for each differentiation pathway is shown. C) Construction of composite networks by combining RT and transcriptional networks. RT networks were used to define the "base network" which include all nodes of RT correlated genes, interaction edges between RT nodes were extracted from TRNs. Composite networks included all RT edges (black undirected lines) transcriptional edges within RT network motifs (directed solid red arrows) and transcriptional edges outside RT network motifs (directed dashed grey arrows). Exemplary composite networks for liver (D) and mesothelium differentiation (E) are shown. For visualization purposes, only RT nodes with correlations >0.9 are shown.

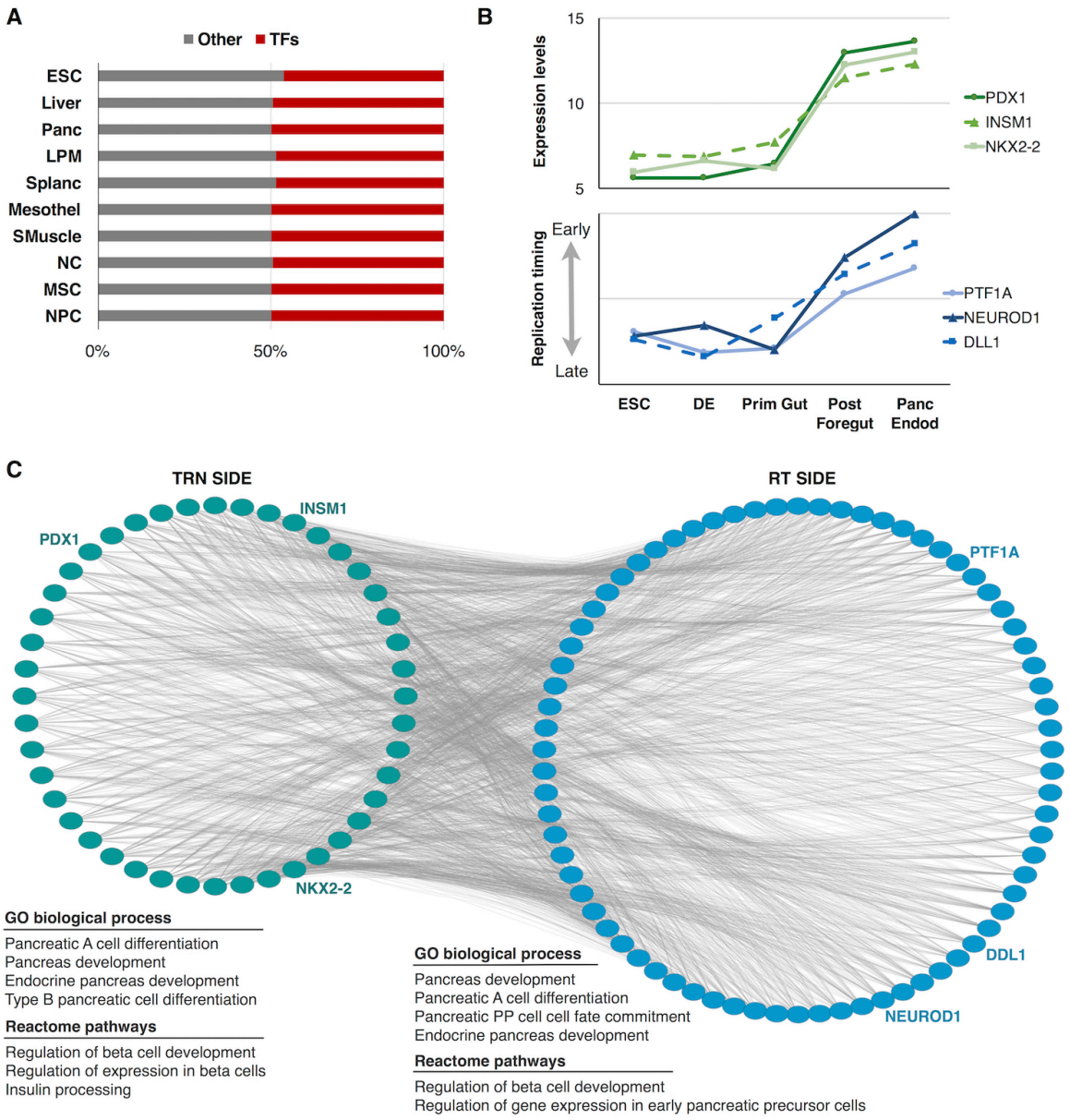


Figure 5: Bipartite networks

A) Transcription factors expression patterns distinguish each cell type/differentiation intermediate. Cell type-specific expression patterns were analyzed to identify the most significant differentially co-expressed genes, then the resulting genes were classified as TFs or other type. B) Expression patterns of exemplary key TFs of pancreas development correlate with the RT of downstream regulators of pancreatic differentiation. C) Bipartite network of pancreas development. The bipartite network was constructed based on the correlation between transcriptional activity of genes in the TRN side and RT changes of genes in the RT side. TRN genes are 50 TFs co-regulated during pancreas development. RT side contains 112 genes with RT patterns highly correlated (>0.8) with the expression levels of genes at the TRN side. Only edges between TRN and RT networks are shown (with each gene in the RT side connected with at least 0.75% of the nodes in the TRN side), as all nodes within each network are connected with all other. Ontology analysis of genes within each network was performed and specific annotations are shown at the bottom.

Methods

Extraction of RT values at the TSS of NCBI RefSeq genes

Replication Timing (RT) data from multiple cell types and intermediate differentiation stages derived from human embryonic stem cells (Rivera-Mulia et al., 2015) were used to extract the RT values at the transcription start sites (TSS) of all RefSeqs genes in R. Briefly, average RT profiles were obtained from replicates and TSS positions were used to predict RT values from the loess smoothed RT profiles (Ryba et al, 2011). This data consists of RT values at the TSS of all RefSeq genes from 15 cell types derived from hESC representing three main germ layers; ectoderm, mesoderm, and endoderm (Figure 1A).

Construction of RT networks based on coordinated changes in RT

We constructed RT networks for different subsets of cell types. We particularly focused on the three major germ layers and the entire set of cell types in our analysis. In our network models, each node represents a gene, and each edge represents a relationship between co-regulated RT of the corresponding two genes in those subset of cell types. More specifically, let us denote the set of genes selected with G = {g_1, g_2, ..., g_n}. Assuming that the number of cell types for the germ layer under consideration (ectoderm, mesoderm, endoderm, or all germ layers) is s. Let us denote the set of cell types with {c_1, c_2, . . . , c_s}. Each gene g_i $\in$ G defines a vector, denoted with w_i. The jth entry of w_i is the replication timing of g_i in cell type c_j. We model the RT network using graph notation as G = (V, E), where V and E denote the set of nodes and edges respectively. Here $\forall$i, 1 $\leq$ i $\leq$ n, node vi $\in$ V corresponds to gene g_i. To construct an edge between two nodes v_i1 and v_i2 of this network, we need to make three key decisions; (i) Is there an edge between v_i1 and v_i2?; (ii) Is gene g_i a switching gene? (iii) What is the physical proximity of the two corresponding genes g_i1 and g_i2 on the chromosome?

For all pairs of genes g_i1, g_i2 $\in$ G, we compute the Pearson's correlation coefficient between their vectors w_i1 and w_i2. If the positive value of this correlation is greater than or equal to a user specified threshold ε then we say that g_i1 and g_i2 are correlated. We draw an edge between v_i1 and v_i2 and insert it to E if g_i1 and g_i2 are correlated.

We consider genes in two categories. (i) We call a gene switching if it is early replicating in at least one cell type and late replicating in at least one other cell type. (ii) We call a gene non-switching otherwise (i.e., if it is consistently early replicating or consistently late replicating in all cell types). A gene having the same (or very similar) RT values across all cell types (non-switching) will yield high correlation values with other genes with such behavior regardless of their RT values. Such a phenomenon will lead into false positive correlations. Thus, we compute the Pearson's correlation coefficient for only the switching gene pairs to avoid false positive correlations. It is worth noting that this is an aggressive filter as some of the non-switching genes may have high variation in their RT values although it is always early or late in replication.

If two genes are located close to each other on the same chromosome, their RT values are expected to be highly correlated due to a natural outcome of the DNA replication process; when the replication starts at a site, it proceeds to the neighboring nucleotides on the chromosome. Such correlations have less significance as compared to those among physically distant genes, for the correlations between distant genes provide hints about the existence of complex interactions that regulate the order in which genes are replicated. To capture this, we classify each edge constructed in our RT network into one of the three categories as follows. For each edge $e \in E$ between nodes v_{i1} and v_{i2} in our graph, we check the locations of g_{i1} and g_{i2} on the DNA. If they are on the same chromosome, we say that they are co-located. Otherwise, we call them not co-located. When the two genes g_{i1} and g_{i2} are co-located, let us denote the difference between their starting positions in the chromosome with $d_{i1,i2}$. Given a user supplied distance threshold (denoted with μ) for the position between two genes, we are now ready to classify edge e into a category:

- Class 1: close. We classify e into this category if g_{i1} and g_{i2} are co-located and $d_{i1,i2} < \mu$.

- Class 2: co-located, distant. We classify e into this category if g_{i1} and g_{i2} are co-located and $d_{i1,i2} \geq \mu$.

- Class 3: not co-located. We classify e into this category if g_{i1} and g_{i2} are not co-located

In Fig1B, if two correlated genes are located as 'b' and 'c', they are considered as close. If one gene among two correlated genes is located as 'a' and the other is located in the range from 'b' to 'c', it is considered as co-located, distant. Finally, if one gene among two correlated genes is located in one chromosome and the other gene is located in a different chromosome, it is considered as not co-located. Thus, we compute the Pearson's correlation coefficient for only not co-located gene pairs to avoid less significant correlations.

From above three decisions we made, we construct RT networks for all-derm, ectoderm, mesoderm, and endoderm. Furthermore, we also construct RT networks for each differentiation path (for example, ESCs -> Lateral plate mesoderm -> Splanchnic mesoderm -> Smooth muscle). We call a gene early replicated if that RT value is greater than 0.3, call a gene late replicated if that RT value is less than -0.3, and call ambiguously replicated if that RT value is between -0.3 and 0.3 (See Fig1E). For each differentiation pathway, we only consider genes that are late replicated in at least one differentiation stage and early replicated in other stage.

RT networks visualization

To visualize our RT network, we focus on a set of highly connected sub networks, called community. First we construct RT network by computing the Pearson's correlation coefficient for only not 'co-located' and 'co-located distant' gene pairs using only RT switching genes. Next, we detect communities in this RT network by running Louvain community detection algorithm (Blondel et al., 2008). This method is a heuristic method that is based on modularity optimization. It is known to outperform all other community detection methods in terms of computation time. Furthermore, the quality of the communities detected is good. After detecting communities in our RT network, all nodes have their own community ID. For clear explanation, we introduce an edge named 'e'. This edge has two nodes, called n_1, and n_2. n_1 has c_1 community ID, and n_2 has c_2 community ID. Let us define d_1 be the number of neighbor nodes that have community ID c_2. Similarly, let us define d_2. We maintain only edges that both d_1 and d_2 is greater than degree_filter_threshold.

With this filtered RT network, we use SAFE algorithm to annotate functional attributes for communities. SAFE (Baryshnikova, 2016) is an automated network annotation algorithm. Given a biological network and a set of functional groups or quantitative features of interest, SAFE performs local enrichment analysis to determine which regions of the network are over-represented for each group of feature. Thus, local neighborhoods were identified and functional

attributes were annotated based on the gene ontology (GO) terms (Baryshnikova, 2016). Finally, we visualize the RT network as 2D maps in Cytoscape by applying force-directed layout algorithm (Shannon et al., 2003) with functionally annotated sub communities.

Construction of Directional RT networks

Directional RT networks were generated for each differentiation pathway for 'late replicated to early replicated' (LtoE) and 'early replicated to late replicated' (EtoL. In our late replicated to early replicated (LtoE) directed RT network model each node represents a gene that switches from late replicated to early replicated (LtoE). In creating these networks, we do not consider how much correlated two corresponding genes are. We only consider causality between two corresponding genes. For differentiation path ESCs (the earliest stage) -> Lateral plate mesoderm -> Splanchnic mesoderm -> Smooth muscle (the latest stage) of late replicated to early replicated (LtoE) directed RT network, we draw a directed edge from a gene that switches in earlier stage to a gene that switches in later stage only if the difference of switching stage is one or two. For example, if LtoE pattern of gene g_1 is L->E->E->E in the differentiation path ESCs -> Lateral plate mesoderm -> Splanchnic mesoderm -> Smooth muscle and LtoE pattern of gene g_2 is L->L->E->E in the same path, we draw an directed edge from g_1 to g_2 because g_1 switches in Lateral plate mesoderm stage (earlier) and g_2 switches in Splanchnic mesoderm stage (later) assuming the change from LtoE of gene g_1 causes the change of gene g_2 in the next stage. Pattern L->E->E->E to pattern L->L->L->E is also valid.

RT network edges overlap with known transcriptional regulatory interactions

We compare the topologies of the RT networks we construct with those of TRNs. Particularly, we use the TRNs constructed using TFs (Neph et al., 2013). To do that, for each cell lineage, we count the number of edges common to its RT network and TRN. Recall that each edge in a RT network may or may not have a direction, whereas all the edges in TRNs have directions. In obtaining the number of common edges, we do not take the direction of edge in TRN into consideration. An undirected edge in the RT network overlaps with a directed edge in the TRN if the gene pairs corresponding to an edge are same in the RT network and the TRN. Using the number of common edges in the two networks, we calculate the p-value of the overlap from their hypergeometric distributions. This p-value quantifies the statistical significance of the evidence they share. The closer the p-value is to zero, the more significant the evidence is. Next, we explain how we calculate the p-value. Let us denote the number of nodes (i.e., genes)

in the given TRN with n. The total number of possible edges in the complete graphs with n nodes is $2\times{}_nC_2$ (i.e., the number of permutations of two nodes). Let us denote this number with M. Also, let us denote the number of edges in the TRN with m, the number of edges in the RT network with K, and the number of edges common to the TRN and the RT network with k. Assume that we have a randomly generated network with the same nodes as those in the RT network. Also, assume that this random network has m edges between randomly selected nodes. Let us denote the number of edges common to the given RT network and this randomly generated network using the random variable X. We compute the probability that the number of common edges between the two networks is equal to a given specific value (say i) as $P(X = i) = ({}_KC_i \times {}_{M-K}C_{m-i}) / ({}_MC_m)$.The numerator in this probability mass function (PMF) describes the number of ways to pick exactly i edges from the RT network in m draws from a complete graph, without replacement. The denominator shows the number of alternative network topologies with the same nodes as the RT network, which has m edges. Using this PMF, we calculate the p-value of having more than or equal to k common edges between the RT network and the TRN as $\sum_{i=k}^{k} P(X = i)$. Thus, smaller p-value shows unexpectedly large number of common edges between the two networks.

Construction of Composite RT and gene expression networks

The composite network merges the interactions observed in RT networks (using only genes present in Neph et al., 2013 networks) with the interactions observed in Transcriptional Regulatory Networks (TRNs). Furthermore, we also visualized the composite networks using the following steps. Base network is a RT network. We then constructed a reduced RT network by considering all possible connected two, three, and four nodes subnetworks from base network. Among such two, three and four nodes RT subnetworks, we exclude non-connected ones. If an edge in the original TRN appears in these all possible subnetworks, then this edge is appended to the reduced TRN network with directed edge. We combine base network and reduced TRN network by taking the union of their edge sets. Additionally, if an edge in the original TRN appears in outside of all possible connected two, three, and four nodes RT subnetworks, then this edge is appended to the composite network with dashed edge.

RT networks motif identification

Motifs are defined as recurrent and statistically significant sub-graphs or patterns. We try to find motifs in the composite networks using the following motif finding algorithm. First, we

created all possible shapes of connected (in terms of undirected edges) two, three, and four nodes subnetworks that are consisted of undirected RT edges and/or directed TRN edges. Then, while iterating all possible subnetworks of composite network, we simply count the number of matching between each subnetwork of composite network and previously created each of all possible shapes. Also, we create multiple shuffled networks that have the same number of nodes and edges with the composite network to differentiate which shape can be a motif. We set z-score as 2.54 for a subnetwork of composite network to be a motif.

Construction of bipartite RT and transcriptional networks

A bipartite graph is a graph with two components. Each component is a set of nodes. In our model, first component is based on the expression patterns of genes, and the second component is based in the replication timing of genes. For our analysis, we started with the list of genes co-expressed in specific cell types that constitute the first component. We append an edge between the first component and the second component if replication timing of a gene in the second component is correlated with expression of a gene in the first component with more than a certain correlation threshold. Next, we removed a gene in the second component if the number of edges of this gene is less than the total number of genes in the first component * 'ratio'. In this way, we generated the list of genes that constitute the second component.

Supplemental Information

Replication Timing Networks: a novel class of gene regulatory networks

Juan Carlos Rivera-Mulia[1], Sebo Kim[2], Haitham Gabr[2], Tamer Kahveci[2,*] and David M. Gilbert[1,3,*]

[1]Department of Biological Science, Florida State University, Tallahassee, FL, 32306-4295, USA.

[1]Department of Computer and Information Sciences and Engineering, University of Florida, Gainesville, Florida 32611, USA.

[3]Center for Genomics and Personalized Medicine, Florida State University, Tallahassee, FL, USA.

	all differentiation pathways	ectoderm	mesoderm	endoderm
RT network size	96	853	623	204
TRN size	3303	843	865	292
Common edges	15	108	71	9
RT network-specific edges	81	745	552	195
TRN-specific edges	3288	735	794	283
Hypergeometric *p-value*	0.04502	4.619e-12	2.815e-07	0.32358

Table S1. Overlap analysis of RT and TRNs interaction edges, related to main Figure 4.

RT networks were constructed for matching cell types in the TRNs (Neph, et al., 2013) and common and unique interaction edges were identified. Only genes within the TRNs were used (475 transcription factors). Hypergeometric test was performed to test the overlap significance (*p-values* are shown). Ectoderm cell types = neural crest, mesenchymal stem cells and neural precursor cells. Mesoderm cell types = lateral plate mesoderm, splanchnic mesoderm, mesothelium and smooth muscle. Endoderm = definitive endoderm, immature hepatic, hepatoblast, liver (hepatocytes), primitive gut, posterior foregut and pancreas (pancreatic endoderm).

Cell type	GO biological process	# genes	Fold Enrichment	P value
ESC	somatic stem cell population maintenance	7	25.03	1.43E-04
	stem cell population maintenance	9	15.5	7.31E-05
DE	cell fate commitment of primary germ layer	5	38.27	2.22E-03
	formation of primary germ layer	7	13.16	1.07E-02
	endodermal cell differentiation	5	26.79	1.26E-02
	endoderm formation	6	25.72	1.37E-03
	endoderm development	6	17.38	1.31E-02
Liver	lipid homeostasis	7	13.74	8.11E-03
	liver development	8	13.44	1.55E-03
	hepaticobiliary system development	8	13.13	1.85E-03
Pancreas	pancreatic A cell differentiation	3	> 100	2.99E-02
	endocrine pancreas development	10	54.95	5.06E-11
	endocrine system development	11	19.32	1.54E-07
	pancreas development	14	41.67	6.64E-15
Mesothel	enteroendocrine cell differentiation	4	45.12	1.94E-02
	glandular epithelial cell differentiation	5	28.2	9.84E-03
	mesodermal cell fate specification	1	30.31	3.25E-02
	mesoderm formation	2	6.33	4.03E-02
	mesoderm morphogenesis	2	6.15	4.24E-02
	mesoderm development	3	5.17	2.08E-02
Smooth Muscle	metanephric smooth muscle tissue development	1	> 100	9.20E-03
	kidney smooth muscle tissue development	1	> 100	9.20E-03
	smooth muscle tissue development	2	21.65	3.98E-03
	muscle tissue development	6	4.53	2.23E-03
	myoblast fate commitment	1	43.3	2.28E-02
	muscle structure development	7	3.34	5.22E-03
NPC	generation of neurons	22	3.3	5.06E-03
	neurogenesis	22	3.08	1.54E-02
NC	neural crest cell migration	5	27.82	1.05E-02
	neural crest cell differentiation	6	18.17	1.02E-02
MSC	positive regulation of mesenchymal cell	3	22.5	3.48E-04
	regulation of mesenchymal cell proliferation	3	17.5	7.19E-04
	mesenchymal cell development	4	13.13	2.68E-04

Table S2. GO analysis of co-expressed genes in each cell type, related to main Figure 5.

Co-expressed genes were identified by weighted correlation network analysis (Langfelder and Horvath, 2008) and ontology analysis (Ashburner et al., 2000; The Gene Ontology Consortium, 2015) using the top 100 genes was performed for each cell type.

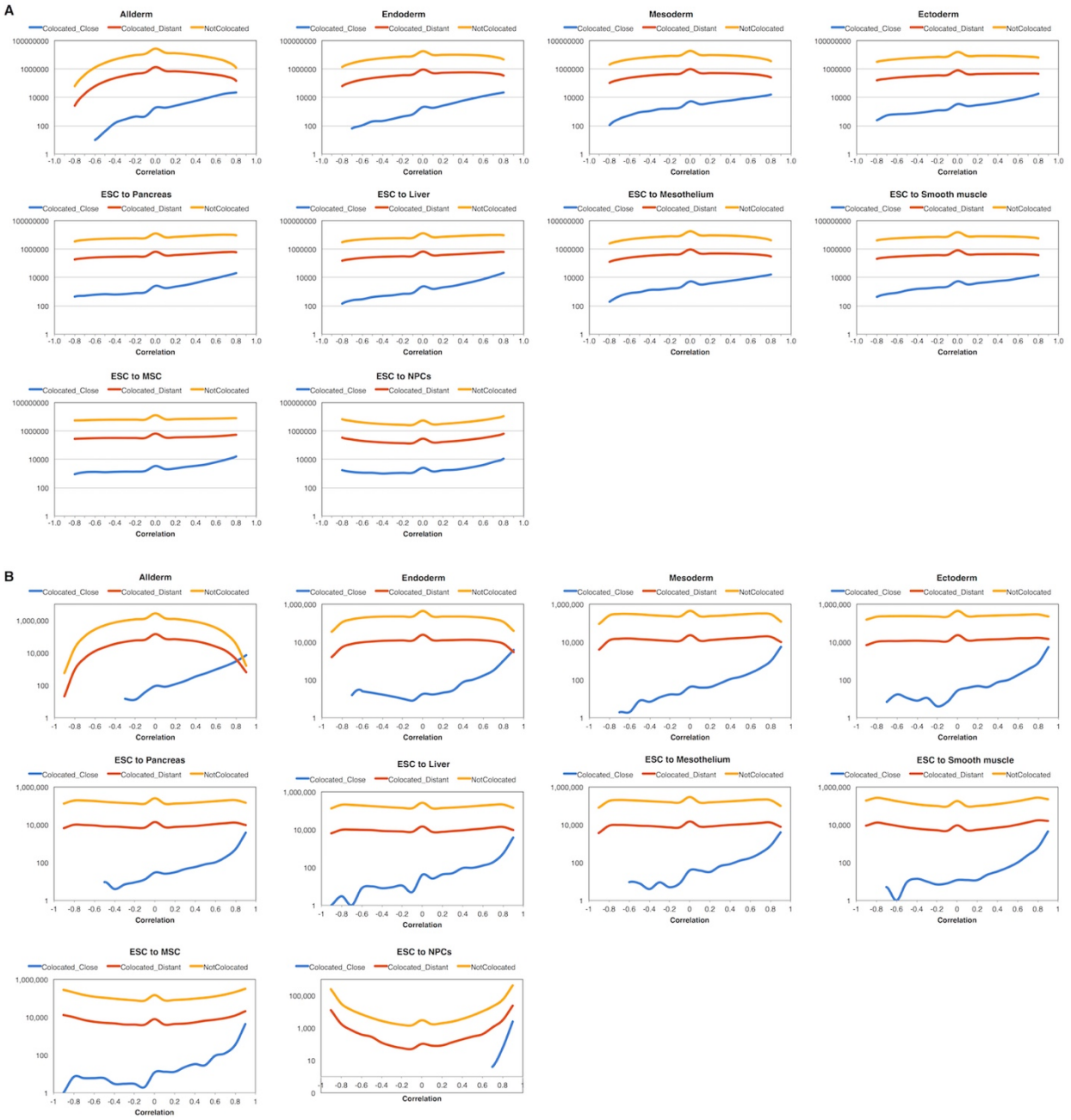


Figure S1. RT correlation per gene pairs, related to main Figure 1.

Number of gene pairs as function of RT correlation for distinct categories of gene pairs: co-located close (within 500kb), co-located distant (separated by > 500kb) and not co-located (from different chromosomes). All gene pairs were computed in (A) and gene pairs between genes that change RT significantly within each differentiation pathway (B) are shown. Only gene pairs with RT correlations >0.75 and located > 500kb apart were considered for RT networks construction.

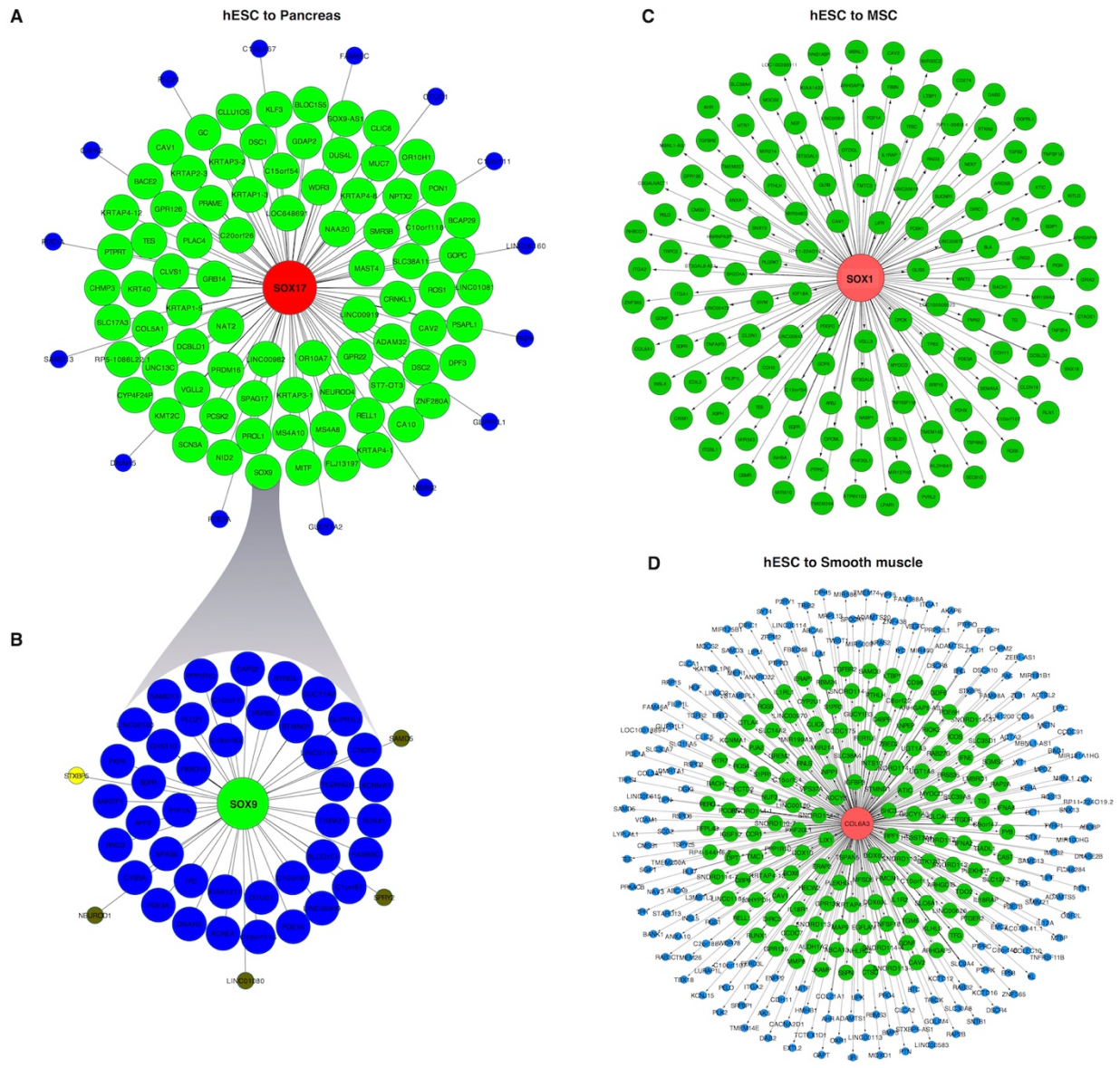


Figure S2. Directional RT networks, related to main Figure 3.

A) An exemplary directional RT network for pancreas differentiation is shown. The central node is SOX17 and all downstream nodes were connected based on times during differentiation at which they change RT. B) A SOX17 downstream network based on the SOX9 gene exemplify downstream relationships in RT networks. C and D) Exemplary directional RT networks for mesenchymal stem cells (MSC) and smooth muscle differentiation.

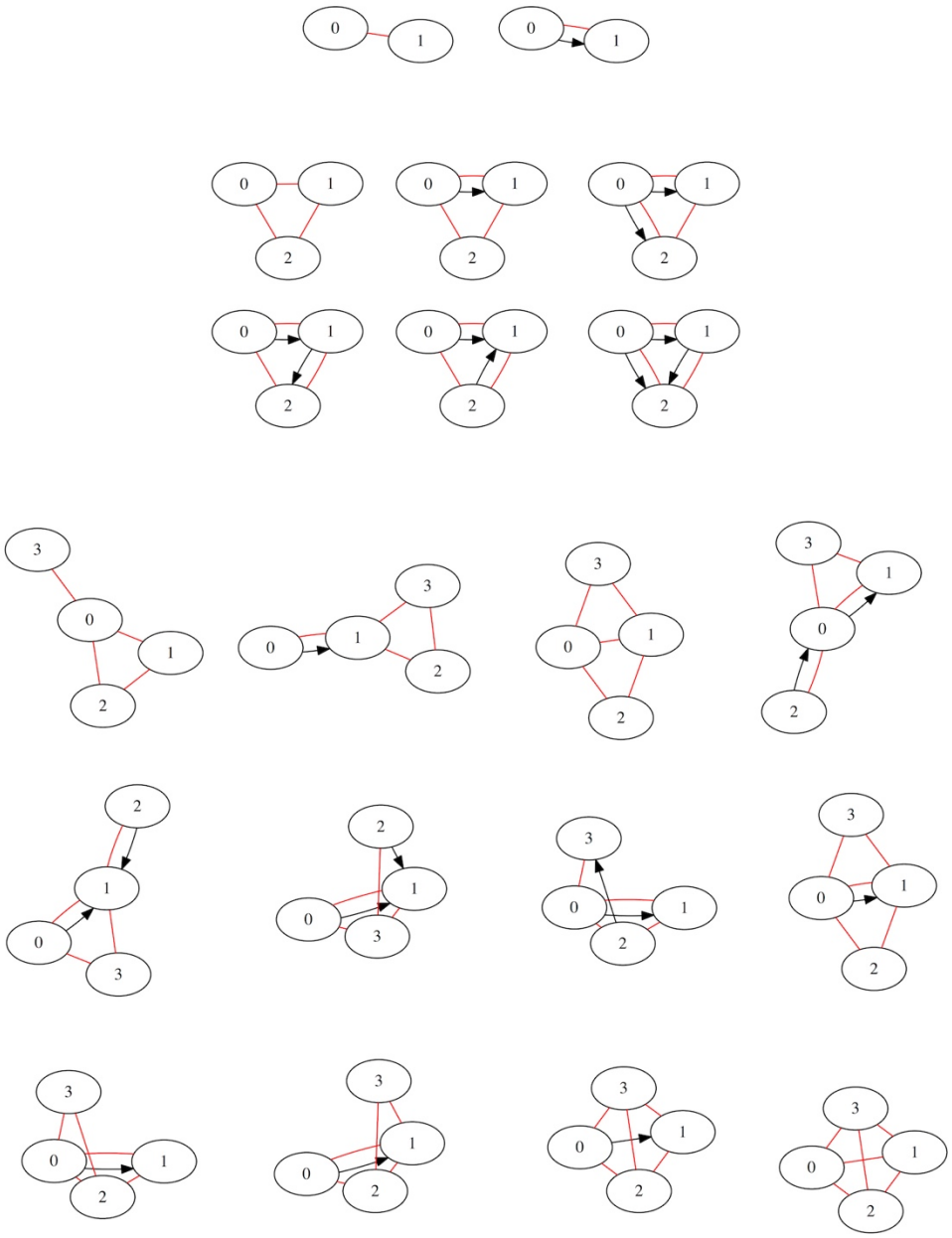


Figure S3. Highly represented motifs identified in liver RT networks, related to main Figure 4.

All possible motifs composed by 2-4 nodes were computed and high occurrence motifs were identified. Statistical significance of each motif pattern was calculated by comparison to randomized networks. Shown are the most frequent motifs of 2-4 nodes in liver RT networks. RT edges are shown in red (undirected edges) and TRN edges are shown in black (directed edges).

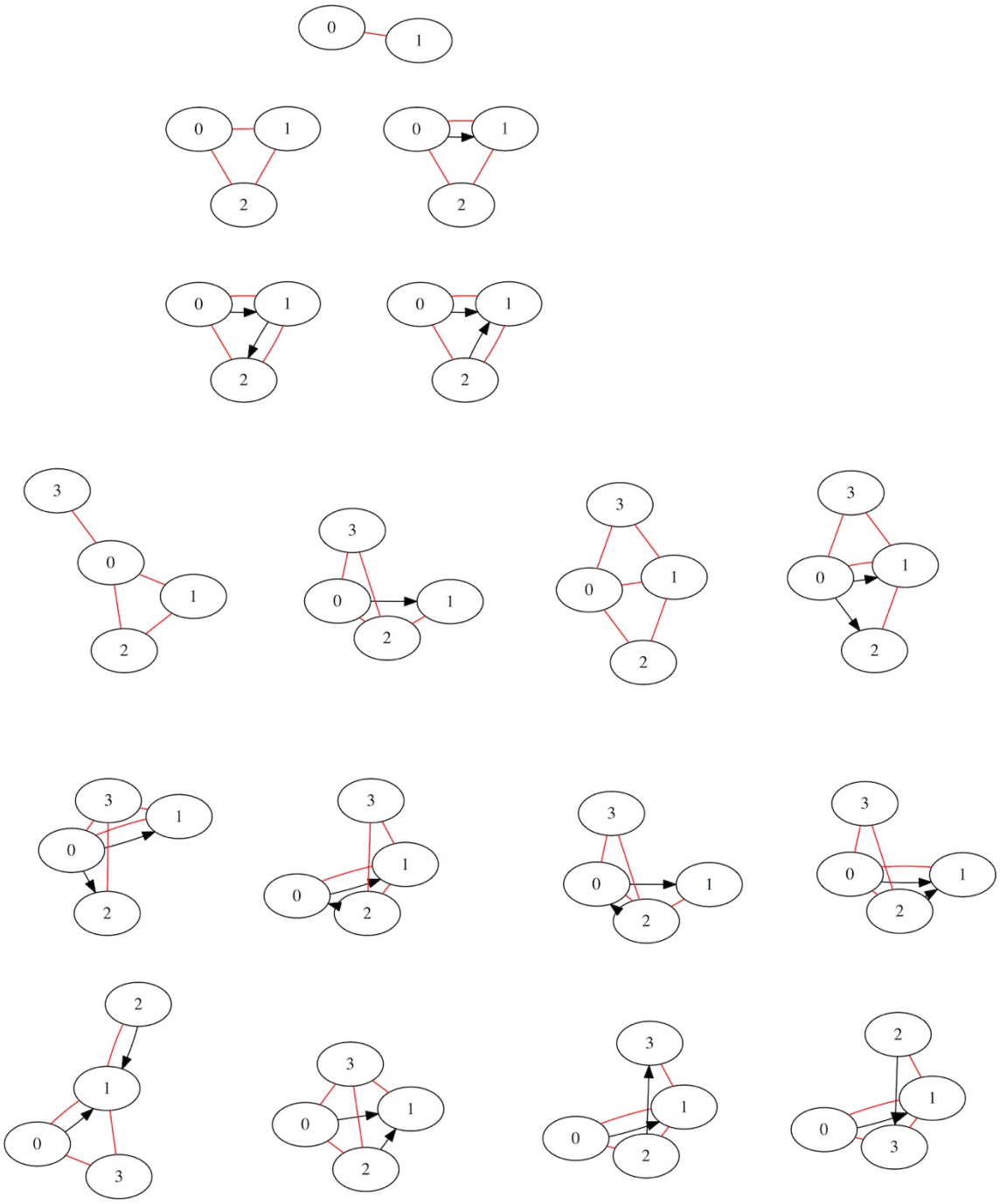


Figure S4. Highly represented motifs identified in mesothelium RT networks, related to main Figure 4.

Most frequent motifs of 2-4 nodes in mesothelium RT networks were identified as in Figure S2. RT edges are shown in red (undirected edges) and TRN edges are shown in black (directed edges).

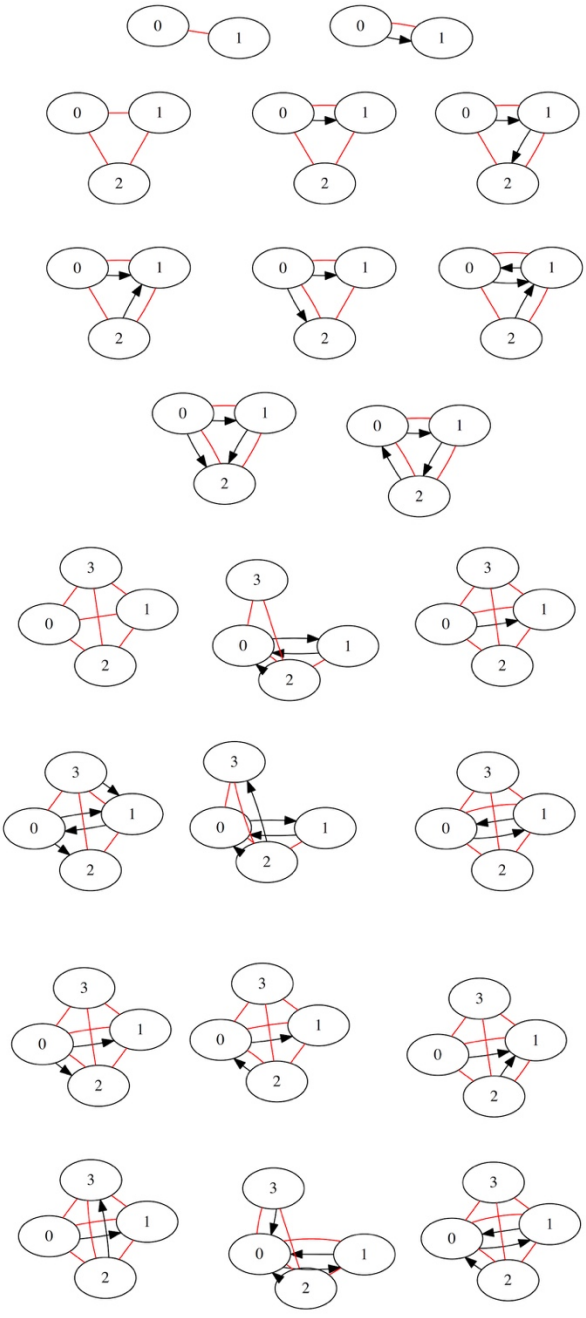


Figure S5. Highly represented motifs identified in NPC RT networks, related to main Figure 4.

Most frequent motifs of 2-4 nodes in NPC RT networks were identified as in Figure S2. RT edges are shown in red (undirected edges) and TRN edges are shown in black (directed edges).